

# 100 years of controversy over standards: an enduring problem

**Gill Elliott** Head of Comparability Programme, Assessment Research & Development

Why are we so bothered about comparability in public examinations? The issue has been a thorn in the sides of educational commentators for at least a century and, despite numerous attempts to solve it, remains a stubborn problem.

This article introduces some of the key issues within the field of comparability, and provides an historical perspective on some of the current concerns. It traces major developments in the theory, methodology and use of comparability research and looks at the way in which theories of comparability have developed and different viewpoints have emerged.

In 1911 a Consultative Committee was convened by the Board of Education to report on *Examinations in Secondary Schools*. What were the comparability-related issues, and how were those issues described and defined?

The 1911 report contained a list of the functions which examinations, in their widest sense, were expected to fulfil at that time. They were expected to:

- test the ability of the candidate for admission to practice a profession;
- ascertain the relative intellectual position of candidates for academic distinction (e.g. scholarships);
- be used for recruitment to the public (civil) service;
- test the efficiency of teachers;
- diffuse a prescribed ideal of liberal<sup>1</sup> culture ("for the efficient discharge of the duties of citizenship in the more responsible positions in life, or as the necessary qualification for admission to University, every young man should be required to have reached a prescribed standard of all round attainment in liberal studies").

This list is still relevant, given the importance of understanding the purposes to which the results of assessments are put when interpreting claims of comparability.

The report outlined the large number of organisations that were providing examinations for the purposes of matriculation and/or progression into the professions. These included not only universities, but also trade organisations and professional societies, such as the London Chamber of Commerce, the Pharmaceutical Society, the Institution of Civil Engineers and the Incorporated Society of Accountants and Auditors. Whilst the many organisations that required examinations still wanted to preserve their own examinations, in 1911 there was beginning to be a move towards recognition of other examinations as equivalent to their own. The document described a system of equivalents being in place, whereby some organisations were prepared to accept alternate examinations of a corresponding standard to their own. However, as the document went on to report, the system was dogged by innumerable confusing restrictions imposed by the various organisations. The main

consequence of the restrictions placed upon the system of equivalents was that the students' choices became very complicated, with an increasing chance of making a poor choice of examination. The document describes it thus:

*While candidates can obtain their Oxford Senior Certificate by passing in five subjects, no one set of five subjects is accepted by all the exempting bodies. A candidate would have to pass in eleven subjects, viz., Arithmetic, English, Mathematics Higher Geometry, Latin, Greek, English History, Geography, French or German, Chemistry or Physics, and a portion of New Testament in Greek, to be sure that his certificate would be accepted by all the bodies who accept the Oxford Senior Certificate as qualifying a candidate for exemption from their Matriculation or Preliminary Examination. If he only passed in the five subjects required by one particular body, and then for any reason changed his plans... he might find it quite useless to him...*  
(Examinations in Secondary Schools, p.34)

Furthermore, a number of awards simply were not accepted as equivalent:

*There are at the present moment a large number of external examinations in Secondary Schools, the certificates of which, regarded as entrance qualifications to the various Universities and professional careers, cannot be said to be always accepted as yet as valid equivalents.*  
(Examinations in Secondary Schools, p.38)

Additionally, there was the difficulty of students who had not yet decided upon a career path and needed a more general qualification, which did not exist. Generally these students took two or even three of the available certificates in order to prepare for a variety of paths. However, the document suggested that this approach might have been slightly unfair, in that it gave these students the option to use the best of their performances.

In 1911 the problem of providing access to the examination to the less able students whilst adequately testing the more able was firmly on the agenda. However, the committee was optimistic about the ability of the system to accomplish this without compromising comparability.

*The levels of attainment reached by different pupils at any one age will of course always differ widely, and it is not supposed that any one set of examination papers will be equally appropriate for them all. But there should be no insuperable difficulty in arriving at a standard which the average pupil should reach at a stated age, and taking this as the criterion by which alternative examinations should be gauged.*  
(Examinations in Secondary Schools, p.90)

<sup>1</sup> Liberal in this context can be defined as 'general broadening'.

The 1911 committee advocated a closer relationship among awarding bodies, and between awarding bodies and the schools, and that the 'standard' be fixed on purely educational grounds. In expanding on the latter point, the report blamed the isolation of awarding bodies from each other for many of the problems and for the fact that even when schools of similar type were compared, standards from different awarding bodies were found to be different (according to a very broad definition of 'standards').

The 1911 report highlighted a number of comparability issues, in particular the problem of aligning standards among Awarding Bodies and the problem of adequately providing a system which would allow some students to qualify for entrance to Universities and professions and others to attain a more general qualification.

The committee proposed a system to accommodate these needs which incorporated two examinations – the School Certificate ("breadth without specialism") and the Higher School Certificate (less general and geared more to the needs of Universities and certain professions). However, they also considered a situation where the former examination could serve a dual purpose – a certificate of general ability, plus a distinction level if certain conditions were met. The rationale behind this was explained in a Board of Education circular (1914), quoted in Norwood (1943), *Curriculum and Examinations in Secondary Schools*:

(iv) *The standard for a pass will be such as may be expected of pupils of reasonable industry and ordinary intelligence in an efficient Secondary School.*

(v) *If the examination is conducted on the principle of easy papers and a high standard of marking, the difference between the standard for a simple pass and that required for matriculation purposes will not be so great as to prevent the same examination being made to serve, as the present school examinations do, both purposes; and with this object a mark of credit will be assigned to those candidates who, in any specific subject or subjects, attain a standard which would be appreciably higher than that required for a simple pass.*

(Curriculum and Examinations in Secondary Schools, 1943, p.27)

It is interesting to note how succinctly these criteria are described, compared with those of today. It is clear that the 'standard' in 1943 was embedded in the notion of the norm.

The following selection of quotations from the Norwood Report (1943) explain how this system began to fall apart.

First, it proved difficult to meet the two distinct purposes of the examination at the same time. The needs of scholars seeking matriculation took precedence, in practice, over those looking for more general certification of educational attainment.

*Whether there was any chance of these two purposes being achieved simultaneously without one obscuring the other is open to doubt; it is easy to be wise after the event; but the history of the examination has shown that the second purpose rapidly overshadowed the first.*

(Curriculum and Examinations in Secondary Schools, 1943, p.27)

The Higher Certificate began to present problems because it had been based upon assumptions that the numbers of candidates would be small and the link with universities close. These assumptions proved mistaken. According to the Norwood Report the certificate became increasingly popular, and attracted increasing numbers of students. This led to new

courses being added to accommodate the needs of an increasingly diverse body of students. These courses fitted less closely to the original conception of the system where the curriculum was closely linked to needs of students seeking a qualification for matriculation.

*...Yet its very success has tended to bring about its progressive disintegration. Rapidly winning recognition on all hands, the certificate awarded on the examination has gathered more authority and more significance than was ever intended at the outset, till it has become a highly coveted possession to every pupil leaving a Secondary School. As the curricula of schools have widened to meet the needs of a Secondary School population rapidly growing more diverse in ability and range of interests, the original structure of the examination has changed. Subjects have necessarily been multiplied, whether susceptible to external examination or not; rules which were framed to give a unity to the curriculum tested by examination have been relaxed. Secondary education has become too varied to be confined within a rigid scheme; teachers are becoming too enterprising to be hedged in by set syllabuses, and subjects themselves are gaining in independence and resourcefulness.*

(Curriculum and Examinations in Secondary Schools, 1943, p.32)

Nevertheless, the Norwood Report was unequivocal about the continuing importance of comparability:

*...If a test is to carry any weight outside the school, there must be some approximation to uniformity of standard in assessing attainment. The test and the verdict must be objective, and conditions must be equal; there can be no prejudice and no favouritism as between school and school or pupil and pupil. Employers, parents and Professional Bodies need the Certificate; employers ask for a disinterested assessment, and would not be satisfied with a Head Master's certificate; parents look for something which will be a hall-mark of their children, valid wherever in the country they may go.*

(Curriculum and Examinations in Secondary Schools, 1943, p.31)

## Changing use of terminology

Before moving on to discuss how our understanding of comparability has progressed since the 1911 report and the Norwood report, it is important to look at definitions of terms. The 1911 and 1943 reports used three of the key terms used currently, shown below, together with their Concise Oxford Dictionary (COD) (Allen, 1992) definition:

- standards: degree of excellence required for the particular purpose
- equivalence: equal in value, amount or importance
- equate: regard as equal or equivalent

To these we should add several further terms:

- alignment: bring into line, place in a straight line
- comparable: that can be compared; fit to be compared
- examinations
- assessments
- qualifications

Confusingly, comparability research over the years has used the latter three terms almost interchangeably. Partly this is due to the historical background. Originally the term 'examinations' was applied both to the

written papers and the overall award. However, that was when 'examinations' (in the sense of the overall award) comprised entirely written papers. Assessment became a term of use to describe components of awards which were not written – coursework, speaking tests etc – and has now tended to become the preferred term to refer to the overall award. Very strictly defined, 'qualification' means the piece of paper which conveys the award, in the same way that 'certificate' does. However, it is also used as the term for the overall award.

The historical papers discussed so far in this article have tended to refer to 'examinations' as the overarching term for assessments which are part of a system of education and training, leading to further educational or employment opportunities. In the remainder of the article (except where reference is being made to the historical documents), 'qualifications' will be used as the preferred term, as it encompasses a wider variety of assessment practice.

It is important to note that the COD definition of 'standards' includes a qualifier – *for a particular purpose*. This is often lost in debates, media headlines and so on. It is also important to note that 'equivalence' and 'alignment' have different meanings. It is possible for qualifications to be aligned according to their equivalence on one particular aspect but to remain non-aligned on other aspects. For example, the subject of General Studies at A level could be compared with other A level subjects on the basis of the amount of teaching time. A comparison made on the basis of prior attainment of students would give a very different result.

## The evolutionary problem in establishing equivalent standards between qualifications

In 1911 the report recognised clearly that the different purposes to which the results of examinations might be put had a bearing upon comparability. The Norwood report identified a key difficulty, which is that, as qualifications evolve, so the underlying assumptions change – which can affect conceptions of comparability. The situation as it developed from 1911 to 1943 is a perfect illustration of this. In 1911 the problem was that multiple qualifications were being used for very similar purposes and they required a degree of inter-changeability. The solution – a single system, with qualifications being used for multiple purposes – was criticised (in 1943) because the qualification in its more multiply-acceptable form attracted more students, who in turn required a greater variety of courses within the system to accommodate their needs. The comparability solutions provided by the original conception of the system were eroded in the face of these challenges.

Both the 1911 report and the 1943 report provide insights into why comparability is so important in the history of English examining. Three main reasons emerge.

First is the relationship between comparability and validity and reliability. The Norwood Report (p.31) is absolutely clear that "*some approximation to uniformity of standard in assessing attainment*" is desirable (if not essential) for examinations to hold any value or currency beyond the school gates. However, it is worth noting the use of 'approximation', and the suggestion of 'uniformity' rather than equivalence. A key aspect of validity is that the inferences made on the basis of the outcomes of any assessment should be meaningful, useful and appropriate, in the particular set of circumstances that they are used. Reliability, which relates to the stability, consistency and precision of an assessment, is strongly linked to validity, because poor reliability

compromises validity. Comparability is a part of validity, as alternative pathways or routes within assessments which lead to the same outcome, or the use of the outcomes of different assessments to access the same FE or employment opportunities, imply a degree of equivalence between them which must be borne out.

Second is the need to provide students with a meaningful choice of qualifications which are recognised by employers and higher education institutions. In 1911 it was proposed that these qualifications should be "valid wherever in the country they may go". Nowadays we might expand this to "wherever in the world they may go". In essence, learners, education institutions and businesses need to be assured of the value of the qualifications.

Third is the social responsibility of awarding bodies to provide students with appropriate qualifications, delivered fairly. In 1943, the Norwood Report referred to an objective test and outcome, taken under equal conditions with no prejudice or favouritism. Today it is expressed in the fact that awarding bodies are committed to ensuring that all assessments are fair, have sound ethical underpinning, and operate according to the highest technical standards.

Having explored in some detail the extent to which educational thinkers early in the twentieth century defined and understood issues of comparability, it is worth tracing briefly some of the more recent developments in theory and practice. For a more detailed description of the evolution of comparability from the mid-nineteenth century to the present, see Tattersall (2007).

Crucial amongst these developments was the move nationally towards measuring, monitoring and maintaining standards between qualifications. This was led primarily by the awarding bodies and regulatory authorities. An unpublished summary (Anonymous, 1970) of early comparability studies recently found in the Archives at Cambridge Assessment reveals that, following discussions at the annual meeting of the Secretaries of GCE examining boards in 1951, it was decided to institute inter-board investigations in a whole series of subjects, at both Ordinary and Advanced level. Nineteen separate studies were described in this paper, investigating inter-Board A level standards from eleven boards including those in England, Wales and N. Ireland. The work encompassed 16 different subjects (and included what may have been the only comparability work ever to have addressed the subjects of Zoology or Botany). These studies were carried out between 1953 and 1968. The report also made reference to similar investigations having been held on subjects at Ordinary Level, but so far no documented evidence of these has come to light. The methods used by the majority of studies carried out in the 1950s and 1960s are familiar to researchers today, as they asked panels of examiner judges, to scrutinise script evidence from key grading points, alongside evidence of demand derived from syllabuses, regulations, 'hurdles' (possibly meaning grade boundaries), and mark schemes. A variety of different judgemental methods of using the script evidence were tried. These included simply reading and discussing the scripts in the light of the demand of papers; re-marking exercises; cross-moderation<sup>2</sup> approaches; and a 'conference' approach. The conference approach involved a review of the practices of the various boards in the subject concerned, and did not incorporate any scrutiny of scripts. Three of the four conferences described related to subject areas already addressed by other forms of comparability study (hence 19 studies in

<sup>2</sup> Cross-moderation methods have been defined as 'systematic ways of looking at candidates' work, that ought to be of the same standard.' (Adams, 2007, p.212)

only 16 subjects). Although the conference approach omitted any investigation of script evidence, it was considered helpful: in the description of the Geography conference (the only subject where a conference approach was taken without any other type of comparability study being conducted), it was stated that:

*This conference brought out yet again the very great value which the investigations and conferences have had over the years in bringing together persons concerned with carrying out similar work for different boards. The interchange of ideas has been valuable and there has undoubtedly been much cross-fertilisation, all of which has contributed towards establishing the comparability of the boards in their demands on candidates and the comparability of the awards made.*

(A review of investigations into subjects at Advanced Level conducted by the GCE boards 1953–1968, p.14.)

Two startling facts about the dedication of the boards to comparability at this time emerge from the summary of comparability studies between 1953 and 1968. In the description of a study carried out in Physics in 1968, the cost of the study is mentioned as being £16,000, which according to two different UK inflation/price conversion tools<sup>3</sup> would equate to about £200,000 today. This was for just one study, albeit one which was designed to test a new method, which included a subject-based reference test taken by a sample of students (the size of the sample was, alas, unrecorded in this summary document) and a questionnaire survey of schools. The second surprising piece of commentary describes the scale of the Mathematics study in 1954:

*There were 20 syllabuses, 50 question papers and nearly 500 scripts; photocopying was not used for the scripts, and the enquiry therefore took three years to complete.*

(A review of investigations into subjects at Advanced Level conducted by the GCE boards 1953–1968, p.5.)

Advances in comparability theory and practice between the 1970s and the present day have been widely and extensively documented. Several reviews were completed of the studies carried out in the 1970s and 1980s (Bardell, Forrest, and Shoemsmith, 1978; Forrest and Shoemsmith, 1985; NEAB, 1996), which largely comprised judgemental cross-moderation approaches. These studies focussed mainly on comparing qualifications on the basis of the perceived demands of the specification and assessment material and/or the perceived quality of examinees' work. As Bramley (2011) has pointed out, both 'perceived demand' and 'perceived quality' might be thought of as higher-order attributes that are built up from lower-order ones and the definition of these attributes suggests that it is appropriate that they be investigated by methods that use the judgement of experts. The development of these methods continued into the 1990s and the use of paired comparisons and Rasch statistical analysis, based upon the work of Louis Thurstone (1959), was added to the research armoury during this period (see Bramley, 2007, for a full history and description of the method). A further refinement to this type of study was the development of a rank-ordering method (Bramley, 2005; Black and Bramley, 2008).

Alongside the development of methods for use with the judgement of experts, alternative statistical methods for assessing the equivalence of

qualifications were explored. These statistical comparisons are based upon different attributes to those used for judgemental comparisons. Attributes for statistical comparisons do not include perceptions of quality or of demand; rather they are based upon some statistical measure applied to a particular population, such as 'percentage gaining grade A', or 'average grade conditional on a given level of prior attainment' (Bramley, 2011). A statistical strand was developed alongside the judgemental method applied to large scale inter-board studies in the 1990s (see Adams *et al.*, 1990 for an early example; also Fowles, 1995; and Pinot de Moira, 2003). Syllabus/subject pairs work has been a feature of research since the early 1970s (Nuttall *et al.* 1974, chapter III) and methods for deriving a 'putative' grade distribution based on prior attainment have been developed more recently.

The final, important, research strand which should be included in this potted history of the development of comparability theory has been the discussions about what is meant by the terms used to define and discuss comparability. Although this has been alluded to throughout the history of comparability (Massey, 1994) it has increased greatly in more recent years, fuelled by debates between individual researchers (Newton, 2005, 2010; Coe, 2007, 2010) and by public events such as the debate: *School exams: what's really happened to 'standards'?*, hosted by Cambridge Assessment on 29th April 2010. The essence of these arguments relates to whether researchers use common terms when discussing comparability, exactly what each term means and how a more common understanding might be brought about. One of the most important recent developments in thinking about comparability is Newton's insight that:

*An issue that has clouded conceptual analysis of comparability in England, perhaps the principal issue, is the failure to distinguish effectively between definitions of comparability and methods for achieving comparability (or methods for monitoring whether comparability has been achieved).* (Newton, 2010, p.288)

## Discussion

It is important to be open and honest about the challenges that are inherent in the study of comparability and assessment processes. Comparability has been an issue for the past century and there are still few completely satisfactory solutions. In this respect an important lesson can be learnt from the 1943 review of the 1911 system changes: if the qualifications are changed, there will be an impact on uptake and use of those qualifications, thus raising further comparability issues. In other words, comparability has always and will always evolve as qualifications do.

In order to go forward, a number of issues need to be addressed:

First, it is important to find clear ways of dealing with the different definitions of comparability, especially when applied to the different purposes to which the results of qualifications are put.

Secondly, Newton (2010) has made it clear that it cannot be assumed that different commentators are talking about the same thing, even when similar terminology is used. There are a number of challenges inherent in the process of explaining comparability evidence to the users of qualifications (students, parents and schools and prospective employers). These include: (i) the confusing nature of the terminology; (ii) the claims which are made both by those organisations delivering qualifications and by wider authoritative bodies (e.g. national and

<sup>3</sup> The currency conversion websites were: <http://safalra.com/other/historical-uk-inflation-price-conversion/> and <http://www.nationalarchives.gov.uk/currency/results.asp#mid>

international government departments and organisations); and (iii) the fact that much of the comparability information that reaches the wider public is conveyed by a third party, such as the media.

Thirdly, it must always be remembered that most of the methods of determining equivalence between qualifications can only ever be accurate to a certain point. A statistical or judgemental method can provide a very specific measure of equivalence, but care must be taken to ensure that it is not spurious, given the statistical limitations of the grades awarded. As Murphy (2010) has stated:

*In measurement terms they [GCSE and A level examinations and the grades which they produce] are 'approximate estimates of educational achievement', which need a great deal of interpretation, rather than precise measurements on a highly sophisticated scale.*

(Murphy, 2010, p.2).

Finally, as qualifications become more high stakes, it needs to be decided whether comparability is the master, or the slave, or neither. The Quality Assurance Agency for Higher Education (2006), stated that:

*...it cannot be assumed that students graduating with the same classified degree from different institutions having studied different subjects, will have achieved similar academic standards; (b) it cannot be assumed that students graduating with the same classified degree from a particular institution having studied different subjects, will have achieved similar academic standards; and (c) it cannot be assumed that students graduating with the same classified degree from different institutions having studied the same subject, will have achieved similar academic standards... These implications are implicitly acknowledged and accepted in the higher education (HE) sector. They are of long standing, and many of those who make use of degree classifications couple this information with their judgement and experience when employing graduates, or recommending awards for further study, or determining salaries. (QAA, 2006, pp.1-2)*

It is important to ensure that the drive for comparability, and the arguments about comparability do not obscure other key aspects of the assessment process, such as fitness for purpose. It is clear from the historical perspective provided in this paper that comparability is an enduring issue, not easily resolved, and that systemic changes inevitably produce further comparability problems. Reviewing the history of these can help to anticipate what may happen in future if changes are made.

## References

- Adams, R. (2007). Cross-moderation methods. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Adams, R.M., Phillips, E.J. & Walker, N.A. (1990). GCSE intergroup comparability study 1989: Music. Organised by the Welsh Joint Education Committee and the Northern Ireland Schools Examination Council for the GCSE and the Joint Council of the GCSE. Available in: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Allen, R.E. (1992). *The concise Oxford dictionary of current English*. 8th Edition. Oxford: Clarendon Press.
- Anonymous (1970). *A review of investigations into subjects at Advanced Level conducted by the GCE boards 1953-1968*. Available from Cambridge Assessment archives.
- Bardell, G. S., Forrest, G. M., & Shoesmith, D. J. (1978). *Comparability in GCE: A review of the boards' studies, 1964-1977*. Manchester: Joint Matriculation Board on behalf of the GCE Examining Boards.
- Black, B. & Bramley, T. (2008). Investigating a judgemental rank-ordering method for maintaining standards in UK examinations. *Research Papers in Education*, **23**, 3, 357-373.
- Board of Education Consultative Committee (1911). Report of the Consultative Committee on examinations in secondary schools. Parliamentary Papers. Session 1911, Vol. XVI, 159. Chairman: Acland, A. H. Dyke.
- Bramley, T. (2005). A rank-ordering method for equating tests by expert judgment. *Journal of Applied Measurement*, **6**, 2, 202-223.
- Bramley, T. (2007). Paired comparison methods. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Bramley, T. (2011). Comparability of examinations standards: Perspectives from Cambridge Assessment Seminar. April 6th 2011, Cambridge.
- Coe, R. (2007). Common examinee methods. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Coe, R. (2010). Understanding comparability of examination standards. *Research Papers in Education*, **25**, 3, 271-284.
- Forrest, G. M. & Shoesmith, D. J. (1985). *A second review of GCE comparability studies*. Manchester: Joint Matriculation Board on behalf of the GCE Examining Boards.
- Fowles, D. (1995). A comparability study in Advanced level Physics. A study based on the Summer 1994 and 1990 examinations. Organised by the NEAB on behalf of the Standing Research Advisory Committee of the GCE Boards. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Massey, A. J. (1994). Standards are slippery! *British Journal of Curriculum and Assessment*, **5**, 1, 37-38.
- Murphy, R. (2004). *Grades of Uncertainty: Reviewing the uses and misuses of examination results*. London: Association of Teachers and Lecturers.
- Murphy, R. (2010). *Grades of Uncertainty? Why examination grades will never give us anything more than very approximate answer to questions about changes in standards over time*. Paper presented at 'School exams: what's really happening to 'standards'?: An event held in London on 29th April 2010. Available at: <http://www.cambridgeassessment.org.uk/ca/Viewpoints/Viewpoint?id=132622>, accessed on July 26th 2010.
- NEAB (1996). *Standards in public examinations, 1977 to 1995: A review of the literature*. Conducted for the School Curriculum and Assessment Authority and the Office for Standards in Education by the Northern Examinations and Assessment Authority on behalf of the GCSE and GCE boards. London: School Curriculum and Assessment Agency.
- Newton, P.E. (2005). Examination standards and the limits of linking. *Assessment in Education: Principles, Policy and Practice*, **12**, 2, 105-123.
- Newton, P.E. (2010). Contrasting conceptions of comparability. *Research Papers in Education*, **25**, 30, 285-292.
- Norwood Report (1943). *Curriculum and Examinations in Secondary Schools*. Report of the Committee of the Secondary School Examinations Council appointed by the President of the Board of Education in 1941. London: HMSO.
- Nuttall, D. L, Backhouse, J. K. & Willmott, A. S. (1974). *Comparability of standards between subjects*. Schools Council Examinations Bulletin 29. London: Evans/Methuen.
- Pinot de Moira, A. (2003). An inter-awarding body comparability study. The statistical analysis of results by awarding body for: GCE A level and

AVCE Business; GCE A level Chemistry; GCSE A level Geography; AVCE Health and Social Care. A study based on the Summer 2002 examination and organised by AQA on behalf of the Joint Council for General Qualifications. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.

Quality Assurance Agency for Higher Education (2006). Background Briefing Note: The classification of degree awards. Gloucester: QAA. Available at: [http://www.qaa.ac.uk/education/briefings/classification\\_20sept06.asp](http://www.qaa.ac.uk/education/briefings/classification_20sept06.asp) accessed on 25.3.2011.

Tattersall, K. (2007). A brief history of policies, practices and issues relating to comparability. In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.

Thurstone, L.L. (1959). *The measurement of values*. Chicago, Illinois: University of Chicago Press.